DE GRUYTER OPEN Open Math. 2015; 13: 33–50

Open Mathematics Open Access

Review Article

Augustyn Markiewicz and Simo Puntanen*

All about the ⊥ with its applications in the linear statistical models

Abstract: For an $n \times m$ real matrix $\mathbf{A}$ the matrix $\mathbf{A}^{\perp}$ is defined as a matrix spanning the orthocomplement of the column space of $\mathbf{A}$, when the orthogonality is defined with respect to the standard inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}'\mathbf{y}$. In this paper we collect together various properties of the $\perp$ operation and its applications in linear statistical models. Results covering the more general inner products are also considered. We also provide a rather extensive list of references.

Keywords: Best linear unbiased estimator, Column space, Generalized inverse, Linear statistical model, Orthocomplement, Orthogonal projector

MSC: 15A99, 62H12, 62J05

DOI 10.1515/math-2015-0005
Received December 13, 2013; accepted April 14, 2014.

1 Introduction

Consider a columnwise partitioned matrix $\mathbf{A} = (\mathbf{a}_1 : \ldots : \mathbf{a}_m) \in \mathbb{R}^{n \times m}$ (the set of $n \times m$ matrices with real elements). Then the column space of $\mathbf{A}$ is defined as

$$\mathscr{C}(\mathbf{A}) = \{\, \mathbf{z} \in \mathbb{R}^n : \mathbf{z} = \mathbf{A}\mathbf{t} = \mathbf{a}_1 t_1 + \cdots + \mathbf{a}_m t_m \text{ for some } \mathbf{t} \in \mathbb{R}^m \,\}.$$

The notation $\mathscr{C}(\mathbf{A})^{\perp}$ refers to the orthocomplement of $\mathscr{C}(\mathbf{A})$, i.e., the set of vectors which are orthogonal (with respect to the standard inner product) to every vector of $\mathscr{C}(\mathbf{A})$:

$$\mathscr{C}(\mathbf{A})^{\perp} = \{\, \mathbf{u} \in \mathbb{R}^n : \mathbf{u}'\mathbf{A}\mathbf{t} = 0 \text{ for all } \mathbf{t} \in \mathbb{R}^m \,\}.$$

Thus, because

$$\mathscr{C}(\mathbf{A})^{\perp} = \{\, \mathbf{u} \in \mathbb{R}^n : \mathbf{u}'\mathbf{A}\mathbf{t} = 0 \text{ for all } \mathbf{t} \in \mathbb{R}^m \,\} = \{\, \mathbf{u} \in \mathbb{R}^n : \mathbf{A}'\mathbf{u} = \mathbf{0} \,\},$$

we have

$$\mathscr{C}(\mathbf{A})^{\perp} = \mathscr{N}(\mathbf{A}') = \text{the null space of } \mathbf{A}'.$$

Now $\mathbf{A}^{\perp}$ is defined as a matrix whose column space is $\mathscr{C}(\mathbf{A}^{\perp}) = \mathscr{C}(\mathbf{A})^{\perp} = \mathscr{N}(\mathbf{A}')$. In view of decomposition $\mathbb{R}^n = \mathscr{C}(\mathbf{A}) \oplus \mathscr{C}(\mathbf{A})^{\perp}$, where $\oplus$ refers to the direct sum, the rank of $\mathbf{A}^{\perp}$ is $\operatorname{rank}(\mathbf{A}^{\perp}) = n - \operatorname{rank}(\mathbf{A})$. The set of all matrices $\mathbf{A}^{\perp}$ is denoted as $\{\mathbf{A}^{\perp}\}$ and hence:

$$\mathbf{Z} \in \{\mathbf{A}^{\perp}\} \iff \text{(a) } \mathbf{A}'\mathbf{Z} = \mathbf{0} \quad \text{and} \quad \text{(b) } \operatorname{rank}(\mathbf{Z}) = n - \operatorname{rank}(\mathbf{A}). \tag{1}$$

Augustyn Markiewicz: Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, PL-60637 Poznań, Poland, E-mail: amark@au.poznan.pl
***Corresponding Author: Simo Puntanen:** School of Information Sciences, FI-33014 University of Tampere, Finland, E-mail: simo.puntanen@uta.fi

We immediately observe that $\mathbf{Z} \in \{\mathbf{A}^{\perp}\} \iff \mathbf{A} \in \{\mathbf{Z}^{\perp}\}$. Trivially $\mathbf{A}^{\perp}$ is unique only when $\mathbf{A}$ is a nonsingular square matrix in which case $\mathbf{A}^{\perp} = \mathbf{0}$. Notice that

$$\mathbf{A} \in \mathbb{R}^{n\times m} \implies \mathbf{A}^{\perp} \in \mathbb{R}^{n\times s}, \quad \text{where} \quad s \geq n - \operatorname{rank}(\mathbf{A}).$$

In this paper our purpose is to review various features of the $\perp$ operation, the “perp-operation”, say, and in particular, to present several useful applications related to linear statistical models. Results covering the more general inner products are also considered. We believe that our review provides a useful summary of the $\perp$ operation and thereby increases the insights and appreciation of this, seemingly simple, operation.

2 $\mathbf{A}^{\perp}$ in terms of generalised inverses

The generalised inverses offer a very handy tool for explicit expressions of the $\mathbf{A}^{\perp}$, and in this section we give a short tour into such possibilities. Matrix $\mathbf{G} \in \mathbb{R}^{m\times n}$ is a generalised inverse of $\mathbf{A} \in \mathbb{R}^{n\times m}$ if

$$\mathbf{AGA} = \mathbf{A}\,, \tag{mp1}$$

and it is the Moore–Penrose inverse, denoted as $\mathbf{A}^{+}$, if it also satisfies the following three conditions:

$$\text{(mp2) } \mathbf{GAG} = \mathbf{G}\,, \qquad \text{(mp3) } (\mathbf{AG})' = \mathbf{AG}\,, \qquad \text{(mp4) } (\mathbf{GA})' = \mathbf{GA}\,.$$

If $\mathbf{G}$ satisfies the condition $\mathbf{AGA} = \mathbf{A}$, we may denote $\mathbf{G} = \mathbf{A}^{-}$, or $\mathbf{G} \in \{\mathbf{A}^{-}\}$. As the excellent references for the generalised inverses, see Ben-Israel & Greville [10] and Rao & Mitra [46]. In particular, for more of about the Moore (of Moore & Penrose), see Ben-Israel [11].

It is well known that the nullspace $\mathscr{N}(\mathbf{A})$ can be expressed as

$$\mathscr{N}(\mathbf{A}_{n\times m}) = \mathscr{C}(\mathbf{I}_m - \mathbf{A}^{-}\mathbf{A})\,,$$

where $\mathbf{A}^{-}$ can be any generalized inverse of $\mathbf{A}$. Hence we can express $\mathscr{C}(\mathbf{A})^{\perp}$ in terms of $\mathbf{A}^{-}$:

$$\mathscr{C}(\mathbf{A})^{\perp} = \mathscr{N}(\mathbf{A}') = \mathscr{C}[\mathbf{I}_n - (\mathbf{A}')^{-}\mathbf{A}'] = \mathscr{C}[\mathbf{I}_n - (\mathbf{A}^{-})'\mathbf{A}']\,. \tag{2}$$

The last equality above follows from the fact

$$\{(\mathbf{A}^{-})'\} = \{(\mathbf{A}')^{-}\}\,. \tag{3}$$

Notice that it is a bit questionable to write $(\mathbf{A}^{-})' = (\mathbf{A}')^{-}$ because (3) means the equality between two *sets*. However, for the (unique) Moore–Penrose inverse we always have $(\mathbf{A}^{+})' = (\mathbf{A}')^{+}$. In light of (2), we have, for example, the following choices for $\mathbf{A}^{\perp}$ (recalling that $\mathbf{A} \in \mathbb{R}^{n\times m}$):

$$\mathbf{I}_n - (\mathbf{A}')^{-}\mathbf{A}', \qquad \mathbf{I}_n - (\mathbf{A}^{-})'\mathbf{A}', \qquad \mathbf{I}_n - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}', \tag{4}$$

where we have used the fact $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-} \in \{(\mathbf{A}')^{-}\}$. By replacing $(\mathbf{A}')^{-}$ with $(\mathbf{A}')^{+}$ in (4) and using

$$(\mathbf{A}')^{+} = (\mathbf{A}^{+})', \qquad (\mathbf{A}')^{+}\mathbf{A}' = \mathbf{AA}^{+},$$

we get

$$\mathbf{I}_n - \mathbf{AA}^{+} := \mathbf{I}_n - \mathbf{P}_{\mathbf{A}} := \mathbf{Q}_{\mathbf{A}} \in \{\mathbf{A}^{\perp}\}\,. \tag{5}$$

It can be shown that if $\mathbf{G}$ satisfies the conditions (mp1) and (mp3), i.e., $\mathbf{G} \in \{\mathbf{A}^{-}_{13}\}$ then $\mathbf{AG}$ is unique and thereby $\mathbf{AA}^{-}_{13} = \mathbf{AA}^{+}$, and hence $\mathbf{I}_n - \mathbf{AA}^{-}_{13}$ is one choice for $\mathbf{A}^{\perp}$.

The notations $\mathbf{P}_{\mathbf{A}}$ and $\mathbf{Q}_{\mathbf{A}}$ in (5) refer to the orthogonal projectors onto $\mathscr{C}(\mathbf{A})$ (with respect to the standard inner product) and $\mathscr{C}(\mathbf{A})^{\perp}$, respectively. Matrix $\mathbf{P}$ is defined as the orthogonal projector onto $\mathscr{C}(\mathbf{A})$ if it satisfies the following conditions:

$$\mathbf{P} = \mathbf{P}' = \mathbf{P}^2 \ \text{ and } \ \mathscr{C}(\mathbf{P}) = \mathscr{C}(\mathbf{A}), \tag{6}$$

which can be shown to be equivalent to $\mathbf{P}(\mathbf{A} : \mathbf{A}^{\perp}) = (\mathbf{A} : \mathbf{0})$, where $(\mathbf{A} : \mathbf{A}^{\perp})$ and $(\mathbf{A} : \mathbf{0})$ denote partitioned matrices.

The matrix $\mathbf{P}$ satisfying (6) is unique and can be written as $\mathbf{P_A} = \mathbf{AA}^{+} = \mathbf{AA}^{-}_{13} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$, where the last expression is invariant for any choice of $(\mathbf{A}'\mathbf{A})^{-}$; this follows from Rao & Mitra [46, Lemma 2.2.4], which says that for nonnull $\mathbf{A}$ and $\mathbf{C}$, the matrix product $\mathbf{AB}^{-}\mathbf{C}$ is invariant with respect to the choice of the generalized inverse $\mathbf{B}^{-}$ if and only if $\mathscr{C}(\mathbf{C}) \subset \mathscr{C}(\mathbf{B})$ and $\mathscr{C}(\mathbf{A}') \subset \mathscr{C}(\mathbf{B}')$. Notice that $\mathbf{AA}^{-}$ is not necessarily an orthogonal projector: it is idempotent and it satisfies $\mathscr{C}(\mathbf{AA}^{-}) = \mathscr{C}(\mathbf{A})$ but it is not necessarily symmetric.

Below is a summary of some of the expressions for $\mathbf{A}^{\perp}$ with obvious extensions to $(\mathbf{A}')^{\perp}$ in terms of generalised inverses.

Theorem 2.1. *For $\mathbf{A} \in \mathbb{R}^{n\times m}$, denote $\mathbf{Q_A} = \mathbf{I}_n - \mathbf{AA}^{+} = \mathbf{I}_n - \mathbf{P_A}$. Then*
(a) $\mathbf{I}_n - (\mathbf{A}')^{-}\mathbf{A}' \in \{\mathbf{A}^{\perp}\}$,
(b) $\mathbf{I}_n - (\mathbf{A}^{-})'\mathbf{A}' \in \{\mathbf{A}^{\perp}\}$,
(c) $\mathbf{I}_n - \mathbf{P_A} = \mathbf{Q_A} \in \{\mathbf{A}^{\perp}\}$,
(d) $\mathbf{I}_n - \mathbf{AA}^{-}_{13} \in \{\mathbf{A}^{\perp}\}$,
(e) $\mathbf{I}_m - \mathbf{A}^{-}\mathbf{A} \in \{(\mathbf{A}')^{\perp}\}$,
(f) $\mathbf{I}_m - \mathbf{A}'(\mathbf{AA}')^{-}\mathbf{A} = \mathbf{I}_m - \mathbf{A}^{+}\mathbf{A} = \mathbf{I}_m - \mathbf{P}_{\mathbf{A}'} = \mathbf{Q}_{\mathbf{A}'} \in \{(\mathbf{A}')^{\perp}\}$.

Obviously the orthogonal projector $\mathbf{Q_A} = \mathbf{I}_n - \mathbf{AA}^{+}$ is often a convenient choice for $\mathbf{A}^{\perp}$ because it is symmetric and idempotent.

3 Some specific formulas

Suppose that $\mathbf{Z}$ is a choice for $\mathbf{A}^{\perp}$. Then, for a comformable matrix $\mathbf{B}$, we have

$$\mathbf{ZB} \in \{\mathbf{A}^{\perp}\} \tag{7}$$

whenever $\operatorname{rank}(\mathbf{ZB}) = \operatorname{rank}(\mathbf{Z})$. According to Marsaglia & Styan [34, Cor. 6.2] (see Theorem 4.1 below),

$$\operatorname{rank}(\mathbf{ZB}) = \operatorname{rank}(\mathbf{Z}) - \dim \mathscr{C}(\mathbf{Z}') \cap \mathscr{C}(\mathbf{B})^{\perp},$$

and thereby (7) holds if and only if $\mathscr{C}(\mathbf{Z}') \cap \mathscr{C}(\mathbf{B})^{\perp} = \{\mathbf{0}\}$. Thus we have the following simple result:

Theorem 3.1. *Let $\mathbf{A} \in \mathbb{R}^{n\times m}$, and $\mathbf{B} \in \mathbb{R}^{n\times q}$. Then for any $\mathbf{A}^{\perp} \in \mathbb{R}^{n\times n}$ we have*

$$\mathbf{A}^{\perp}\mathbf{B} \in \{\mathbf{A}^{\perp}\} \iff \mathscr{C}[(\mathbf{A}^{\perp})'] \cap \mathscr{C}(\mathbf{B})^{\perp} = \{\mathbf{0}\}.$$

In particular, choosing $\mathbf{Q_A}$ as $\mathbf{A}^{\perp}$ yields

$$\mathbf{Q_A B} \in \{\mathbf{A}^{\perp}\} \iff \mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathbb{R}^{n},$$

where $(\mathbf{A} : \mathbf{B})$ denotes the partioned $n \times (m+q)$ matrix.

In the next theorem we take a look at the perps of some particular partitioned matrices.

Theorem 3.2. *Let $\mathbf{A} \in \mathbb{R}^{n\times m}$, and $\mathbf{B} \in \mathbb{R}^{n\times q}$. Then for any $\mathbf{A}^{\perp}$ we have*
(a) $\begin{pmatrix} \mathbf{A}^{\perp} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} \in \left\{ \begin{pmatrix} \mathbf{A}_{n\times m} \\ \mathbf{0}_{q\times m} \end{pmatrix}^{\perp} \right\}$,
(b) $\begin{pmatrix} \mathbf{I}_n \\ -\mathbf{B}' \end{pmatrix} \mathbf{A}^{\perp} \in \left\{ \begin{pmatrix} \mathbf{A}_{n\times m} & \mathbf{B}_{n\times q} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix}^{\perp} \right\}$,
(c) $\begin{pmatrix} \mathbf{I}_n \\ -\mathbf{B}' \end{pmatrix} \in \left\{ \begin{pmatrix} \mathbf{B}_{n\times q} \\ \mathbf{I}_q \end{pmatrix}^{\perp} \right\}$.

Proof. Part (a) is obvious as the orthogonality condition corresponding to (1a) trivially holds and

$$\operatorname{rank}\begin{pmatrix}\mathbf{A}^{\perp} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix} = n - \operatorname{rank}(\mathbf{A}) + q.$$

To prove (b), we observe that

$$[(\mathbf{A}^{\perp})' : -(\mathbf{A}^{\perp})'\mathbf{B}]\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix} = \mathbf{0}\,.$$

Moreover, the rank of $\begin{pmatrix}\mathbf{I}_n\\ -\mathbf{B}'\end{pmatrix}\mathbf{A}^{\perp}$ is

$$\operatorname{rank}\big[(\mathbf{A}^{\perp})' : -(\mathbf{A}^{\perp})'\mathbf{B}\big] = \operatorname{rank}[(\mathbf{A}^{\perp})'] = n - \operatorname{rank}(\mathbf{A})\,,$$

while

$$\begin{aligned}\operatorname{rank}\begin{pmatrix}\mathbf{A}_{n\times m} & \mathbf{B}_{n\times q}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix}^{\perp} &= n + q - \operatorname{rank}\begin{pmatrix}\mathbf{A}_{n\times m}\\ \mathbf{0}\end{pmatrix} - \operatorname{rank}\begin{pmatrix}\mathbf{B}_{n\times q}\\ \mathbf{I}_q\end{pmatrix}\\ &= n + q - \operatorname{rank}(\mathbf{A}) - q = n - \operatorname{rank}(\mathbf{A})\,.\end{aligned}$$

Thus (b) is confirmed. Part (c) can be proved in the corresponding way. □

Theorem 3.3. *Consider* $\mathbf{A} \in \mathbb{R}^{n\times m}$ *and* $\mathbf{B} \in \mathbb{R}^{t\times m}$. *Then for any* $\mathbf{A}^{\perp}$ *and* $\mathbf{B}^{\perp}$ *we have*

$$\begin{pmatrix}\mathbf{A}^{\perp} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}^{\perp}\end{pmatrix} \in \left\{\begin{pmatrix}\mathbf{A}\\ \mathbf{B}\end{pmatrix}^{\perp}\right\}$$

if and only if $\mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B}') = \{\mathbf{0}\}$.

Proof. The orthogonality condition (1a) obviously holds while

$$\operatorname{rank}\begin{pmatrix}\mathbf{A}^{\perp} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}^{\perp}\end{pmatrix} = n + t - \operatorname{rank}(\mathbf{A}) - \operatorname{rank}(\mathbf{B}),$$

$$\operatorname{rank}\begin{pmatrix}\mathbf{A}\\ \mathbf{B}\end{pmatrix}^{\perp} = n + t - \operatorname{rank}(\mathbf{A}) - \operatorname{rank}(\mathbf{B}) + \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B}').$$

Thus the proof is completed. □

Remark 3.4. *It might be a bit tempting to rewrite part (a) of Theorem 3.2 as*

$$\begin{pmatrix}\mathbf{A}_{n\times m}\\ \mathbf{0}_{q\times m}\end{pmatrix}^{\perp} = \begin{pmatrix}\mathbf{A}^{\perp} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix}. \tag{8}$$

However, expression like (8) *is obviously problematic, and the meaning of the above notation should be clarified. One interpretation for* (8) *might be to agree that it means that*

$$\left\{\begin{pmatrix}\mathbf{A}_{n\times m}\\ \mathbf{0}_{q\times m}\end{pmatrix}^{\perp}\right\} = \left\{\begin{pmatrix}\mathbf{A}^{\perp} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix}\right\}. \tag{9}$$

In other words, the sets of matrices are identical. However, the statement (9) *is incorrect as can be concluded by Theorem 3.5 below.* □

Let us ask the following: which matrices $\mathbf{B} \in \mathbb{R}^{n\times p}$ and $\mathbf{D} \in \mathbb{R}^{q\times p}$ satisfy the following:

$$\begin{pmatrix}\mathbf{A}^{\perp} & \mathbf{B}_{n\times p}\\ \mathbf{0} & \mathbf{D}_{q\times p}\end{pmatrix} \in \left\{\begin{pmatrix}\mathbf{A}_{n\times m}\\ \mathbf{0}_{q\times m}\end{pmatrix}^{\perp}\right\}\,?$$

We first observe that equation

$$\begin{pmatrix} (\mathbf{A}^{\perp})' & \mathbf{0} \\ \mathbf{B}' & \mathbf{D}' \end{pmatrix} \begin{pmatrix} \mathbf{A}_{n\times m} \\ \mathbf{0}_{q\times m} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{n\times m} \\ \mathbf{0}_{q\times m} \end{pmatrix}$$

holds if and only if $\mathscr{C}(\mathbf{B}) \subset \mathscr{C}(\mathbf{A})^{\perp}$. Supposing that $\mathscr{C}(\mathbf{B}) \subset \mathscr{C}(\mathbf{A})^{\perp}$, then, in view of Marsaglia & Styan [34, Cor. 19.1], the rank of $\left(\begin{smallmatrix} \mathbf{A}^{\perp} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{smallmatrix}\right)$ is additive on the Schur complement, i.e.,

$$\operatorname{rank}\begin{pmatrix} \mathbf{A}^{\perp} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} = \operatorname{rank}(\mathbf{D}) + \operatorname{rank}(\mathbf{A}^{\perp} - \mathbf{B}\mathbf{D}^{-}\mathbf{0}) = \operatorname{rank}(\mathbf{D}) + \operatorname{rank}(\mathbf{A}^{\perp}) = \operatorname{rank}(\mathbf{D}) + n - \operatorname{rank}(\mathbf{A})\,.$$

On the other hand, because

$$\operatorname{rank}\begin{pmatrix} \mathbf{A}_{n\times m} \\ \mathbf{0}_{q\times m} \end{pmatrix}^{\perp} = n + q - \operatorname{rank}(\mathbf{A})\,,$$

we immediately obtain the following:

Theorem 3.5. *Consider $\mathbf{A} \in \mathbb{R}^{n\times m}$, $\mathbf{B} \in \mathbb{R}^{n\times p}$, and $\mathbf{D} \in \mathbb{R}^{q\times p}$. Then the relation*

$$\begin{pmatrix} \mathbf{A}^{\perp} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \in \left\{ \begin{pmatrix} \mathbf{A} \\ \mathbf{0} \end{pmatrix}^{\perp} \right\}$$

holds if and only if $\mathscr{C}(\mathbf{B}) \subset \mathscr{C}(\mathbf{A})^{\perp}$ and $\mathbf{D}$ has full row rank.

4 Two rank formulas and a decomposition of orthogonal projector

Two particular rank formulas in terms of the orthocomplement are worth special praising due to their numerous applications particularly when dealing with linear statistical models: the rank of the product $\mathbf{A}_{n\times a}\mathbf{B}_{a\times m}$ and the rank of the partitioned matrix $(\mathbf{A}_{n\times a} : \mathbf{B}_{n\times b})$.

Theorem 4.1. *The rank of the partitioned matrix $(\mathbf{A}_{n\times a} : \mathbf{B}_{n\times b})$ can be expressed as*

$$\operatorname{rank}(\mathbf{A} : \mathbf{B}) = \operatorname{rank}(\mathbf{A}) + \operatorname{rank}[(\mathbf{A}^{\perp})'\mathbf{B}] = \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B}'\mathbf{A}^{\perp})\,, \tag{10}$$

and the rank of the matrix product $\mathbf{A}_{n\times a}\mathbf{B}_{a\times m}$ is

$$\operatorname{rank}(\mathbf{A}\mathbf{B}) = \operatorname{rank}(\mathbf{A}) - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B})^{\perp}. \tag{11}$$

In terms of an arbitrary generalized inverse $\mathbf{A}^{-}$, (10) can be expressed as

$$\operatorname{rank}(\mathbf{A} : \mathbf{B}) = \operatorname{rank}[\mathbf{A} : (\mathbf{I}_n - \mathbf{A}\mathbf{A}^{-})\mathbf{B}] = \operatorname{rank}(\mathbf{A}) + \operatorname{rank}[(\mathbf{I}_n - \mathbf{A}\mathbf{A}^{-})\mathbf{B}] = \operatorname{rank}(\mathbf{A}) + \operatorname{rank}[(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{B}]\,. \tag{12}$$

As a reference to (10) and (12) we may mention Marsaglia & Styan [34, Th. 5]. For the references to (11), see, e.g., Marsaglia & Styan [34, Cor. 6.2], Rao [43, p. 28], and Zyskind & Martin [59, p. 1194]. We may also mention that O.M. Baksalary & Trenkler [8] provide several expressions for the ranks of a product of two matrices and of a column-wise partitioned matrix as well as an extensive list of related references.

Several applications of (10) and (11) appear in Puntanen, Styan & Isotalo [39, Ch. 5]. One example concerns the decomposition of the column space $\mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp})$, where $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\mathbf{V}$ is an $n \times n$ (symmetric) nonnegative definite matrix. Such a situation occurs when we consider the general linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{denoted as } \mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}, \tag{13}$$

where $\mathbf{X}$ is a known $n \times p$ model matrix, the vector $\mathbf{y}$ is an observable n-dimensional random vector, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is an unobservable vector of random errors with expectation $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, and covariance matrix $\operatorname{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}$. Then we have the following; see, e.g., Rao [45, Lemma 2.1].

Theorem 4.2. *Consider the linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, defined as in* (13). *Then*

$$\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) .$$

Moreover, if the model is correct, in which case it is called consistent, then the observed (realized) value of the random vector $\mathbf{y}$ *satisfies*

$$\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) . \tag{14}$$

For a discussion concerning the consistency concept, see, e.g., Puntanen & Styan [38], J.K. Baksalary, Rao & Markiewicz [5], Groß [16, p. 314], and Tian et al. [53]. In this paper, we assume that the corresponding consistency holds whatever model we have.

When working with linear models, we often need to consider the orthogonal projector onto the column space of the partitioned matrix. Then the following theorem appears to be very convenient in various connections; see, e.g., Puntanen, Styan & Isotalo [39, Th. 8] and Seber & Lee [50, Appendix B].

Theorem 4.3. *The orthogonal projector* (*with respect to the standard inner product*) *onto the column space* $\mathscr{C}(\mathbf{A}_{n\times a} : \mathbf{B}_{n\times b})$ *can be decomposed as*

$$\mathbf{P}_{(\mathbf{A}\,:\,\mathbf{B})} = \mathbf{P}_{\mathbf{A}} + \mathbf{P}_{(\mathbf{I}_n - \mathbf{P}_{\mathbf{A}})\mathbf{B}} = \mathbf{P}_{\mathbf{A}} + \mathbf{P}_{\mathscr{C}(\mathbf{A}\,:\,\mathbf{B})\cap\mathscr{C}(\mathbf{A})^{\perp}} . \tag{15}$$

We complete this section by some remarks on the explicit expression for the intersection of $\mathscr{C}(\mathbf{A})$ and $\mathscr{C}(\mathbf{B})$. For a reference, see Rao & Mitra [46, Complement 7, p. 118].

Theorem 4.4. *Consider the matrices* $\mathbf{A}_{n\times a}$ *and* $\mathbf{B}_{n\times b}$ *and denote* $\mathbf{Q}_{\mathbf{B}} = \mathbf{I}_n - \mathbf{P}_{\mathbf{B}}$. *Then*

$$\mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}) = \mathscr{C}\big[\mathbf{A}(\mathbf{A}'\mathbf{B}^{\perp})^{\perp}\big] = \mathscr{C}\big[\mathbf{A}(\mathbf{A}'\mathbf{Q}_{\mathbf{B}})^{\perp}\big] = \mathscr{C}\big[\mathbf{A}(\mathbf{I}_a - \mathbf{P}_{\mathbf{A}'\mathbf{Q}_{\mathbf{B}}})\big] = \mathscr{C}\big(\mathbf{A}[\mathbf{I}_a - (\mathbf{A}'\mathbf{Q}_{\mathbf{B}}\mathbf{A})^{-}\mathbf{A}'\mathbf{Q}_{\mathbf{B}}\mathbf{A}]\big) .$$

It is obvious that

$$\mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B})^{\perp} = \mathscr{C}\big[\mathbf{A}(\mathbf{A}'\mathbf{B})^{\perp}\big] = \mathscr{C}\big[\mathbf{A}(\mathbf{I}_a - \mathbf{P}_{\mathbf{A}'\mathbf{B}})\big] .$$

In particular, if $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\mathbf{V}_{n\times n}$ is nonnegative definite, then

$$\mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{V})^{\perp} = \mathscr{C}\big[\mathbf{X}(\mathbf{X}'\mathbf{V})^{\perp}\big] = \mathscr{C}\big[\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{\perp}\big] = \mathscr{C}\big(\mathbf{X}[\mathbf{I}_p - (\mathbf{X}'\mathbf{V}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}\mathbf{X}]\big) ,$$

and in view of $\mathscr{C}(\mathbf{M}) \cap \mathscr{C}(\mathbf{V})^{\perp} = \mathscr{C}(\mathbf{X} : \mathbf{V})^{\perp}$,

$$\mathbf{M}[\mathbf{I}_n - (\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{M}] \in \big\{(\mathbf{X} : \mathbf{V})^{\perp}\big\},$$

where $\mathbf{M} = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}}$. Notice also that according to Theorem 4.3 we have $\mathbf{P}_{(\mathbf{X}:\mathbf{V})} = \mathbf{P}_{\mathbf{X}} + \mathbf{P}_{\mathbf{M}\mathbf{V}}$ and thereby

$$\mathbf{I}_n - \mathbf{P}_{(\mathbf{X}:\mathbf{V})} = \mathbf{M} - \mathbf{P}_{\mathbf{M}\mathbf{V}} = \mathbf{M}(\mathbf{I}_n - \mathbf{P}_{\mathbf{M}\mathbf{V}}) \in \big\{(\mathbf{X} : \mathbf{V})^{\perp}\big\} .$$

5 Orthocomplement when the inner product matrix is $\mathbf{V}$

5.1 $\mathbf{V}$ is positive definite

Consider now the inner product in $\mathbb{R}^n$ defined as $\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbf{V}} = \mathbf{x}'\mathbf{V}\mathbf{y}$, where $\mathbf{V}$ is a positive definite symmetric matrix. The orthocomplement of $\mathscr{C}(\mathbf{A}_{n\times m})$ with respect to this inner product is

$$\mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}} = \{\, \mathbf{y} \in \mathbb{R}^n : \mathbf{z}'\mathbf{A}'\mathbf{V}\mathbf{y} = 0 \text{ for all } \mathbf{z} \in \mathbb{R}^m \,\} .$$

By $\mathbf{A}^{\perp}_{\mathbf{V}}$ we will denote any matrix whose column space is $\mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}}$. Recall that $\mathbf{A}^{\perp}_{\mathbf{I}}$ is shortly denoted as $\mathbf{A}^{\perp}$. We have

$$\mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}} = \{\, \mathbf{y} \in \mathbb{R}^n : \mathbf{A}'\mathbf{V}\mathbf{y} = \mathbf{0} \,\} = \mathscr{N}(\mathbf{A}'\mathbf{V}) = \mathscr{C}(\mathbf{V}\mathbf{A})^{\perp} = \mathscr{C}(\mathbf{V}^{-1}\mathbf{A}^{\perp}) ,$$

where the last equality can be concluded from

$$\mathbf{A}'\mathbf{V}\cdot\mathbf{V}^{-1}\mathbf{A}^{\perp} = \mathbf{0} \implies \mathscr{C}(\mathbf{V}^{-1}\mathbf{A}^{\perp}) \subset \mathscr{N}(\mathbf{A}'\mathbf{V}),$$

and

$$\operatorname{rank}(\mathbf{V}^{-1}\mathbf{A}^{\perp}) = \operatorname{rank}(\mathbf{A}^{\perp}) = n - \operatorname{rank}(\mathbf{A}) = \dim\mathscr{C}(\mathbf{V}\mathbf{A})^{\perp}.$$

Notice that corresponding to (1),

$$\mathbf{Z} \in \{\mathbf{A}^{\perp}_{\mathbf{V}}\} \iff \text{(a) } \mathbf{A}'\mathbf{V}\mathbf{Z} = \mathbf{0} \text{ and (b) } \operatorname{rank}(\mathbf{Z}) = n - \operatorname{rank}(\mathbf{A}).$$

Remark 5.1. *Obviously we can write* $\mathbf{V}^{-1}\mathbf{A}^{\perp} \in \{\mathbf{A}^{\perp}_{\mathbf{V}}\}$ *and* $\mathbf{V}^{-1}\mathbf{A}^{\perp} \in \{(\mathbf{V}\mathbf{A})^{\perp}\}$. *Question: Is it correct to write*

$$\{\mathbf{A}^{\perp}_{\mathbf{V}}\} = \{(\mathbf{V}\mathbf{A})^{\perp}\}\ ?$$

It is easy to confirm that the answer is positive. □

Now we have the following decomposition:

$$\mathbb{R}^n = \mathscr{C}(\mathbf{A}) \oplus \mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}} = \mathscr{C}(\mathbf{A}) \oplus \mathscr{C}(\mathbf{V}^{-1}\mathbf{A}^{\perp}),$$

and hence every $\mathbf{y} \in \mathbb{R}^n$ has a unique representation as a sum

$$\mathbf{y} = \mathbf{A}\mathbf{b} + \mathbf{V}^{-1}\mathbf{A}^{\perp}\mathbf{c} = \mathbf{y}_* + \dot{\mathbf{y}},$$

for some $\mathbf{b}$ and $\mathbf{c}$. The vector $\mathbf{y}_* = \mathbf{A}\mathbf{b}$ is the orthogonal projection of $\mathbf{y}$ onto $\mathscr{C}(\mathbf{A})$ along $\mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}}$. The orthogonal projector $\mathbf{P}_{\mathbf{A};\mathbf{V}}$ is such a matrix which transforms $\mathbf{y}$ into its projection $\mathbf{y}_*$, i.e., $\mathbf{P}_{\mathbf{A};\mathbf{V}}\mathbf{y} = \mathbf{y}_* = \mathbf{A}\mathbf{b}$. Its explicit unique representation is

$$\mathbf{P}_{\mathbf{A};\mathbf{V}} = \mathbf{A}(\mathbf{A}'\mathbf{V}\mathbf{A})^{-}\mathbf{A}'\mathbf{V}.$$

We may mention that part (a) of Theorem 3.2 holds even if the inner product matrix is $\mathbf{V}$, i.e.,

$$\begin{pmatrix} \mathbf{A}^{\perp}_{\mathbf{V}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} \in \left\{ \begin{pmatrix} \mathbf{A}_{n\times m} \\ \mathbf{0}_{q\times m} \end{pmatrix}^{\perp}_{\mathbf{V}} \right\}.$$

Similarly Theorems 3.3 and 3.5 hold also when all orthocomplements are taken with respect to the inner product matrix $\mathbf{V}$.

5.2 V is nonnegative definite, possibly singular

Let $\mathbf{V}$ be a singular nonnegative definite matrix. Then $\langle \mathbf{t}, \mathbf{u} \rangle_{\mathbf{V}} = \mathbf{t}'\mathbf{V}\mathbf{u}$ is a semi-inner product and the corresponding seminorm (squared) is $\|\mathbf{t}\|^2_{\mathbf{V}} = \mathbf{t}'\mathbf{V}\mathbf{t}$. For a singular nonnegative definite matrix $\mathbf{V}$ we can define the matrix $\mathbf{A}^{\perp}_{\mathbf{V}}$ again as any matrix spanning $\mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}}$, and so

$$\mathscr{C}(\mathbf{A}^{\perp}_{\mathbf{V}}) = \mathscr{C}(\mathbf{A})^{\perp}_{\mathbf{V}} = \mathscr{N}(\mathbf{A}'\mathbf{V}) = \mathscr{C}(\mathbf{V}\mathbf{A})^{\perp}.$$

As noted by Puntanen, Styan & Isotalo [39, §2.5] for (even) a singular $\mathbf{V}$ we do have the decomposition

$$\mathbb{R}^n = \mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{A}^{\perp}_{\mathbf{V}}) = \mathscr{C}(\mathbf{A}) + \mathscr{C}(\mathbf{V}\mathbf{A})^{\perp}, \tag{16}$$

but, however, the above decomposition is not necessarily a direct sum. For any nonnegative definite $\mathbf{V}$ we have, on account of Theorem 4.1,

$$\dim\mathscr{C}(\mathbf{V}\mathbf{A})^{\perp} = n - \operatorname{rank}(\mathbf{V}\mathbf{A}) = [n - \operatorname{rank}(\mathbf{A})] + \dim\mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{V})^{\perp},$$

which means that (16) becomes a direct sum decomposition if and only if $\mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{V})^{\perp} = \{\mathbf{0}\}$.

For the characterization of the generalized orthogonal projector, see Mitra & Rao [36]. Some related considerations appear also in Harville [19, §14.12.i], Rao & Rao [47, p. 81], and Tian & Takane [54, 55].

5.3 Some further considerations

Consider the linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, defined as in (13), and let $\mathbf{V}$ be positive definite. Then we have observed that the following sets are identical:

$$\begin{array}{lll} \text{(a)}\ \mathscr{C}(\mathbf{X})^{\perp}_{\mathbf{V}^{-1}}\,, & \text{(b)}\ \mathscr{C}(\mathbf{V}\mathbf{X}^{\perp})\,, & \text{(c)}\ \mathscr{N}(\mathbf{X}'\mathbf{V}^{-1})\,, \\ \text{(d)}\ \mathscr{C}(\mathbf{V}^{-1}\mathbf{X})^{\perp}, & \text{(e)}\ \mathscr{N}(\mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}})\,, & \text{(f)}\ \mathscr{C}(\mathbf{I}_n - \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}})\,. \end{array}$$

For (a), ..., (f) above, see also Puntanen, Styan & Isotalo [39, §5.13]. When $\mathbf{V}$ is singular, the above considerations become more complicated. A very convenient tool appears to be the following class of matrices:

$$\mathcal{W} = \{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\}\,. \tag{17}$$

In (17) $\mathbf{U}$ can be any $p \times p$ matrix as long as $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$ is satisfied. Of course, $\mathbf{U}$ can be chosen as $\mathbf{0}$ if $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$ which happens, for example, when $\mathbf{V}$ is positive definite. The set $\mathcal{W}$ of matrices has an important role in the theory of linear models. Below are listed some useful equivalent statements concerning $\mathbf{W}$:

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W})\,, \tag{18a}$$

$$\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{W})\,, \tag{18b}$$

$$\mathbf{X}'\mathbf{W}^{-}\mathbf{X} \text{ is invariant for any choice of } \mathbf{W}^{-}, \tag{18c}$$

$$\mathscr{C}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X}) = \mathscr{C}(\mathbf{X}') \text{ for any choice of } \mathbf{W}^{-}, \tag{18d}$$

$$\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{X} = \mathbf{X} \text{ for any choices of the generalized inverses involved.} \tag{18e}$$

Moreover, each of these statements is equivalent to $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}')$, and hence to the statements (18b')–(18e') obtained from (18b)–(18e), by setting $\mathbf{W}'$ in place of $\mathbf{W}$. As the references to (18), we may mention J.K. Baksalary, Puntanen & Styan [4, Th. 2], J.K. Baksalary & Mathew [3, Th. 2], and Harville [19, p. 468].

According to Puntanen, Styan & Isotalo [39, §5.13] the following now holds.

Theorem 5.2. *Suppose that* $\mathbf{X}$ *is an* $n \times p$ *matrix,* $\mathbf{V}$ *is an* $n \times n$ *nonnegative definite matrix and* $\mathbf{W} \in \mathcal{W}$, *where* $\mathcal{W}$ *is defined as in* (17). *Then*

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{W}^{-}\mathbf{X} : \mathbf{I}_n - \mathbf{W}^{-}\mathbf{W})^{\perp},$$

where $\mathbf{W}^{-}$ *is an arbitrary (but fixed) generalized inverse of* $\mathbf{W}$. *The column space* $\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp})$ *can be expressed also as*

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}\big[(\mathbf{W}^{-})'\mathbf{X} : \mathbf{I}_n - (\mathbf{W}^{-})'\mathbf{W}'\big]^{\perp}.$$

Moreover, let $\mathbf{V}$ *be possibly singular and assume that* $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$. *Then*

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{V}^{-}\mathbf{X} : \mathbf{I}_n - \mathbf{V}^{-}\mathbf{V})^{\perp} \subset \mathscr{C}(\mathbf{V}^{-}\mathbf{X})^{\perp},$$

where the inclusion becomes equality if and only if $\mathbf{V}$ *is positive definite.*

Remark 5.3. *It is of interest to note that the perp symbol* $\perp$ *drops down, so to say, very "nicely" when* $\mathbf{V}$ *is positive definite:*

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp})^{\perp} = \mathscr{C}(\mathbf{V}^{-1}\mathbf{X})\,,$$

but when $\mathbf{V}$ *is singular we have to use a much more complicated rule to drop down the* $\perp$ *symbol:*

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp})^{\perp} = \mathscr{C}(\mathbf{W}^{-}\mathbf{X} : \mathbf{I}_n - \mathbf{W}^{-}\mathbf{W})\,,$$

where $\mathbf{W} \in \mathcal{W}$. □

Remark 5.4. *Let us next prove the following: If* $\mathbf{W} \in \mathcal{W}$, *where* $\mathcal{W}$ *is defined as in* (17), *then*

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{W}^{-}\mathbf{X})^{\perp} \iff \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^{n}. \tag{19}$$

We first observe that

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{W}^{-}\mathbf{X} : \mathbf{I}_n - \mathbf{W}^{-}\mathbf{W})^{\perp} = \mathscr{C}(\mathbf{W}^{-}\mathbf{X})^{\perp} \cap \mathscr{C}(\mathbf{I}_n - \mathbf{W}^{-}\mathbf{W})^{\perp} .$$

Thus we always have $\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) \subset \mathscr{C}(\mathbf{W}^{-}\mathbf{X})^{\perp}$, *where the equality appears only if* $\dim \mathscr{C}(\mathbf{W}^{-}\mathbf{X})^{\perp} = \operatorname{rank}(\mathbf{V}\mathbf{X}^{\perp})$. *Now we have*

$$\begin{aligned} \operatorname{rank}(\mathbf{V}\mathbf{X}^{\perp}) &= \operatorname{rank}(\mathbf{W}) - \operatorname{rank}(\mathbf{X}) , \\ \dim \mathscr{C}(\mathbf{W}^{-}\mathbf{X})^{\perp} &= n - \operatorname{rank}(\mathbf{W}^{-}\mathbf{X}) = n - \operatorname{rank}(\mathbf{X}) , \end{aligned}$$

from which our claim (19) *follows.* □

For completeness we state the following related result, due to Rao & Mitra [46, p. 140].

Theorem 5.5. *Consider the linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ *and denote* $\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{X}'$, *where* $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$, *and let* $\mathbf{W}^{-}$ *be an arbitrary generalized inverse of* $\mathbf{W}$. *Then*

$$\begin{aligned} \mathscr{C}(\mathbf{W}^{-}\mathbf{X}) \oplus \mathscr{C}(\mathbf{X})^{\perp} &= \mathbb{R}^n, \qquad & \mathscr{C}(\mathbf{W}^{-}\mathbf{X})^{\perp} \oplus \mathscr{C}(\mathbf{X}) &= \mathbb{R}^n, \\ \mathscr{C}[(\mathbf{W}^{-})'\mathbf{X}] \oplus \mathscr{C}(\mathbf{X})^{\perp} &= \mathbb{R}^n, \qquad & \mathscr{C}[(\mathbf{W}^{-})'\mathbf{X}]^{\perp} \oplus \mathscr{C}(\mathbf{X}) &= \mathbb{R}^n. \end{aligned}$$

6 Statistical examples

6.1 Centering

We would like to start with a simple but at the same time very important orthocomplement in statistics: the set of vectors orthogonal to the vector of ones, that is, $\mathscr{C}(\mathbf{1}_n)^{\perp}$, where $\mathbf{1}_n = (1, 1, \dots, 1)' \in \mathbb{R}^n$. In what follows, we most of the time drop off the subscript from the vector $\mathbf{1}_n$; from the context its dimension should be obvious. The orthogonal projector onto $\mathscr{C}(\mathbf{1}_n)$ is $\mathbf{P}_{\mathbf{1}} = \frac{1}{n}\mathbf{1}\mathbf{1}' := \mathbf{J}$ and the orthogonal projector onto $\mathscr{C}(\mathbf{1}_n)^{\perp}$ is $\mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}' := \mathbf{C}$; $\mathbf{C}$ is the centering matrix.

Consider the $n \times 2$ data matrix $\mathbf{U}$ partitioned as

$$\mathbf{U} = (\mathbf{x} : \mathbf{y}) = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_{(1)}' \\ \mathbf{u}_{(2)}' \\ \vdots \\ \mathbf{u}_{(n)}' \end{pmatrix} .$$

Here $\mathbf{u}_{(i)} = \binom{x_i}{y_i} \in \mathbb{R}^2$ represents the ith case or the ith observation in the observation space, and the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ represent the two variables in the variable space. Let $\bar{\mathbf{u}} = \binom{\bar{x}}{\bar{y}} \in \mathbb{R}^2$ denote the mean vector of x- and y-variables and $\mathbf{S}$ the sample covariance matrix:

$$\begin{aligned} \bar{\mathbf{u}} &= \tfrac{1}{n}\mathbf{U}'\mathbf{1}_n = \tfrac{1}{n}(\mathbf{u}_{(1)} + \mathbf{u}_{(2)} + \dots + \mathbf{u}_{(n)}) = \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} , \\ \mathbf{S} &= \tfrac{1}{n-1}\mathbf{U}'\mathbf{C}\mathbf{U} = \tfrac{1}{n-1}\sum_{i=1}^{n}(\mathbf{u}_{(i)} - \bar{\mathbf{u}})(\mathbf{u}_{(i)} - \bar{\mathbf{u}})'. \end{aligned}$$

Now the following theorem is easy to confirm; for details, see, e.g., Puntanen, Styan & Isotalo [39, Ch. 3].

Theorem 6.1. *For conformable matrices, the following statements hold:*

(a) *The vector* $\bar{\bar{\mathbf{y}}} = \bar{y}\mathbf{1}$ *is the orthogonal projection of the variable vector* $\mathbf{y}$ *onto the column space* $\mathscr{C}(\mathbf{1})$: $\bar{\bar{\mathbf{y}}} = \bar{y}\mathbf{1} = \mathbf{J}\mathbf{y} = \mathbf{P}_{\mathbf{1}}\mathbf{y}$.

(b) *The centered variable vector* $\tilde{\mathbf{y}}$ *is the orthogonal projection of* $\mathbf{y}$ *onto the column space* $\mathscr{C}(\mathbf{1})^{\perp}$: $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{J}\mathbf{y} = \mathbf{C}\mathbf{y} = (\mathbf{I}_n - \mathbf{P}_{\mathbf{1}})\mathbf{y}$.

(c) *Let the variances of the variables x and y be nonzero, i.e.,* $\mathbf{x} \notin \mathscr{C}(\mathbf{1})$ *and* $\mathbf{y} \notin \mathscr{C}(\mathbf{1})$. *Then the sample correlation coefficient r_{xy} is the cosine of the angle between the centered variable vectors:*

$$r_{xy} = \cos(\mathbf{Cx}, \mathbf{Cy}) = \cos(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \frac{\mathbf{x}'\mathbf{Cy}}{\sqrt{\mathbf{x}'\mathbf{Cx} \cdot \mathbf{y}'\mathbf{Cy}}} .$$

(d) $\mathbf{y}$ *is centered* $\iff \mathbf{y} \in \mathscr{C}(\mathbf{1})^{\perp} = \mathscr{C}(\mathbf{C}) = \mathscr{N}(\mathbf{1}')$.

Next we shortly consider a typical $n \times p$ model matrix $\mathbf{X}$ partitioned as $\mathbf{X} = (\mathbf{1} : \mathbf{x}_1 : \ldots : \mathbf{x}_k) = (\mathbf{1} : \mathbf{X}_0)$, and so $p = k+1$. The sample covariance matrix of the x-variables is $\mathbf{S}_{\mathbf{xx}} = \frac{1}{n-1}\mathbf{X}_0'\mathbf{C}\mathbf{X}_0$ and the sample correlation matrix is $\mathbf{R}_{\mathbf{xx}} = [\operatorname{diag}(\mathbf{S}_{\mathbf{xx}})]^{-1/2}\mathbf{S}_{\mathbf{xx}}[\operatorname{diag}(\mathbf{S}_{\mathbf{xx}})]^{-1/2}$. While calculating the correlations, we assume that all x-variables have nonzero variances, that is, the matrix $\operatorname{diag}(\mathbf{T}_{\mathbf{xx}})$ is positive definite, or in other words: $\mathbf{x}_i \notin \mathscr{C}(\mathbf{1})$, $i = 1, \ldots, k$. Theorem 4.1 implies then the following result:

Theorem 6.2. *The rank of the model matrix* $\mathbf{X} = (\mathbf{1} : \mathbf{X}_0)$ *can be expressed as*

$$\operatorname{rank}(\mathbf{X}) = 1 + \operatorname{rank}(\mathbf{X}_0) - \dim \mathscr{C}(\mathbf{1}) \cap \mathscr{C}(\mathbf{X}_0) = \operatorname{rank}(\mathbf{1} : \mathbf{C}\mathbf{X}_0) = 1 + \operatorname{rank}(\mathbf{C}\mathbf{X}_0) = 1 + \operatorname{rank}(\mathbf{S}_{\mathbf{xx}}) ,$$

and thereby

$$\operatorname{rank}(\mathbf{S}_{\mathbf{xx}}) = \operatorname{rank}(\mathbf{X}) - 1 = \operatorname{rank}(\mathbf{C}\mathbf{X}_0) = \operatorname{rank}(\mathbf{X}_0) - \dim \mathscr{C}(\mathbf{1}) \cap \mathscr{C}(\mathbf{X}_0) .$$

If all x-variables have nonzero variances, i.e., the correlation matrix $\mathbf{R}_{\mathbf{xx}}$ *is properly defined, then* $\operatorname{rank}(\mathbf{R}_{\mathbf{xx}}) = \operatorname{rank}(\mathbf{S}_{\mathbf{xx}})$. *Moreover, the following statements are equivalent:*

$$\text{(a)} \det(\mathbf{S}_{\mathbf{xx}}) \neq 0, \qquad \text{(b)} \operatorname{rank}(\mathbf{X}) = k+1, \qquad \text{(c)} \operatorname{rank}(\mathbf{X}_0) = k \text{ and } \mathbf{1} \notin \mathscr{C}(\mathbf{X}_0).$$

For the rank of of the sample covariance matrix, see Trenkler [56]. As regards the geometry and linear models, the reader may take a look at Margolis [31], Herr [22], and Seber [49].

6.2 Estimability in a simple ANOVA

Following Puntanen, Styan & Isotalo [39, §1.2], consider a simple analysis-of-variance (ANOVA) model

$$\mathscr{A}: \quad \mathbf{y} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{1}_{n_2} & \mathbf{0} & \mathbf{1}_{n_2} & \ldots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_g} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{1}_{n_g} \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_g \end{pmatrix} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = (\mathbf{1}_n : \mathbf{X}_0) \begin{pmatrix} \mu \\ \boldsymbol{\tau} \end{pmatrix} + \boldsymbol{\varepsilon} ,$$

where $n = n_1 + \cdots + n_g$. As the rank of the $n \times (g+1)$ model matrix $\mathbf{X}$ is g we know that $\boldsymbol{\beta}$ is not estimable under $\mathscr{A}$. Which parametric functions of $\boldsymbol{\beta}$ are estimable?

We recall that $\mathbf{K}'\boldsymbol{\beta}$ is estimable if it has an unbiased linear estimator, say $\mathbf{Ay}$ with property $\operatorname{E}(\mathbf{Ay}) = \mathbf{AX}\boldsymbol{\beta} = \mathbf{K}'\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$, i.e., $\mathbf{AX} = \mathbf{K}'$. Hence the parametric function $\mathbf{k}'\boldsymbol{\beta}$ is estimable under $\mathscr{A}$ if and only if

$$\mathbf{k} \in \mathscr{C}(\mathbf{X}') = \mathscr{C}\begin{pmatrix} \mathbf{1}_n' \\ \mathbf{X}_0' \end{pmatrix} = \mathscr{C}\begin{pmatrix} \mathbf{1}_g' \\ \mathbf{I}_g \end{pmatrix} . \tag{20}$$

In view of part (c) of Theorem 3.2, one choice for $\begin{pmatrix} \mathbf{1}_g' \\ \mathbf{I}_g \end{pmatrix}^{\perp}$ is $\begin{pmatrix} -1 \\ \mathbf{1}_g \end{pmatrix}$, i.e.,

$$\begin{pmatrix} -1 \\ \mathbf{1}_g \end{pmatrix} := \mathbf{u} \in \left\{ \begin{pmatrix} \mathbf{1}_g' \\ \mathbf{I}_g \end{pmatrix}^{\perp} \right\} .$$

Hence, according to (20), the parametric function $\mathbf{k}'\boldsymbol{\beta}$ is estimable if and only if

$$\mathbf{k} \in \left(\mathscr{C}\begin{pmatrix}\mathbf{1}_g' \\ \mathbf{I}_g\end{pmatrix}^{\perp}\right)^{\perp} = \mathscr{C}(\mathbf{u})^{\perp},$$

i.e.,

$$\mathbf{k}'\mathbf{u} = 0, \quad \text{where} \quad \mathbf{u} = \begin{pmatrix}-1 \\ \mathbf{1}_g\end{pmatrix}. \tag{21}$$

We can also study the estimability of a parametric function of $\tau_1, \dots, \tau_g$ (dropping off the parameter μ); denote this function as $\ell'\boldsymbol{\tau}$. Then

$$(0, \ell')\begin{pmatrix}\mu \\ \boldsymbol{\tau}\end{pmatrix} = \ell'\boldsymbol{\tau}\,,$$

and on account of (21), the estimability condition for $\ell'\boldsymbol{\tau}$ becomes $\ell'\mathbf{1}_g = 0$.

6.3 Best linear unbiased estimator, BLUE

An unbiased linear estimator $\mathbf{Gy}$ for $\mathbf{X}\boldsymbol{\beta}$ is defined to be the best linear unbiased estimator, BLUE, for $\mathbf{X}\boldsymbol{\beta}$ under the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ if $\operatorname{cov}(\mathbf{Gy}) \leq_{\mathrm{L}} \operatorname{cov}(\mathbf{Ly})$ for all $\mathbf{L}\colon \mathbf{LX} = \mathbf{X}$, where "$\leq_{\mathrm{L}}$" refers to the Löwner partial ordering. In other words, $\mathbf{Gy}$ has the smallest covariance matrix in the Löwner sense among all linear unbiased estimators. The following theorem gives the "fundamental BLUE equation"; see, e.g., Rao [40], Zyskind [58], J.K. Baksalary [1], and O.M. Baksalary & Trenkler [6, 7].

Theorem 6.3. *Consider the general linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, defined as in* (13). *Then the estimator $\mathbf{Gy}$ is the* BLUE *for $\mathbf{X}\boldsymbol{\beta}$ if and only if $\mathbf{G}$ satisfies the equation*

$$\mathbf{G}(\mathbf{X} : \mathbf{VX}^{\perp}) = (\mathbf{X} : \mathbf{0})\,. \tag{22}$$

Notice also that even though $\mathbf{G}$ in (22) may not be unique, the numerical observed value of $\mathbf{Gy}$ is unique (with probability 1) once the random vector $\mathbf{y}$ has obtained its value in the space $\mathscr{C}(\mathbf{X} : \mathbf{VX}^{\perp})$. The set of matrices $\mathbf{G}$ satisfying (22) is sometimes denoted as $\{\mathbf{P}_{\mathbf{X}|\mathbf{VX}^{\perp}}\}$.

Remark 6.4. *At this point we may take a liberty to make a short side trip to the notation $\mathbf{P}_{\mathbf{A}|\mathbf{B}}$ in the spirit of Rao [45] and Kala [25]. Supposing that $\mathscr{C}(\mathbf{A})$ and $\mathscr{C}(\mathbf{B})$ are (virtually) disjoint, then $\mathbf{y} \in \mathscr{C}(\mathbf{A} : \mathbf{B})$ has a unique representation as a sum $\mathbf{y} = \mathbf{y}_A + \mathbf{y}_B$, where $\mathbf{y}_A \in \mathscr{C}(\mathbf{A})$, $\mathbf{y}_B \in \mathscr{C}(\mathbf{B})$. A matrix $\mathbf{P}$ which transforms every $\mathbf{y} \in \mathscr{C}(\mathbf{A} : \mathbf{B})$ into its projection $\mathbf{y}_A$ is called a projector onto $\mathscr{C}(\mathbf{A})$ along $\mathscr{C}(\mathbf{B})$. It appears that the projector $\mathbf{P} := \mathbf{P}_{\mathbf{A}|\mathbf{B}}$ onto $\mathscr{C}(\mathbf{A})$ along $\mathscr{C}(\mathbf{B})$ may be defined by the equation*

$$\mathbf{P}_{\mathbf{A}|\mathbf{B}}(\mathbf{A} : \mathbf{B}) = (\mathbf{A} : \mathbf{0})\,.$$

Kala [25, Lemma 2.5] proved that if $\mathscr{C}(\mathbf{A}) \cap \mathscr{C}(\mathbf{B}) = \{\mathbf{0}\} = \mathscr{C}(\mathbf{C}) \cap \mathscr{C}(\mathbf{D})$, then

$$\{\mathbf{P}_{\mathbf{C}|\mathbf{D}}\} \subset \{\mathbf{P}_{\mathbf{A}|\mathbf{B}}\} \quad \iff \quad \mathscr{C}(\mathbf{A}) \subset \mathscr{C}(\mathbf{C}) \text{ and } \mathscr{C}(\mathbf{B}) \subset \mathscr{C}(\mathbf{D})\,.$$

Moreover, Rao [45] showed that

$$(\mathbf{P}_{\mathbf{VA}|\mathbf{A}^{\perp}} + \mathbf{P}_{\mathbf{A}^{\perp}|\mathbf{VA}})\mathbf{z} = \mathbf{z}\,, \quad (\mathbf{P}_{\mathbf{VA}^{\perp}|\mathbf{A}} + \mathbf{P}_{\mathbf{A}|\mathbf{VA}^{\perp}})\mathbf{y} = \mathbf{y}\,, \quad \mathbf{P}_{\mathbf{A}|\mathbf{VA}^{\perp}}\mathbf{y} = (\mathbf{I}_n - \mathbf{P}'_{\mathbf{A}^{\perp};\mathbf{V}})\mathbf{y}\,,$$

hold for all $\mathbf{z} \in \mathscr{C}(\mathbf{A}^{\perp} : \mathbf{VA}) = \mathscr{C}(\mathbf{A}^{\perp} : \mathbf{V})$ and $\mathbf{y} \in \mathscr{C}(\mathbf{A} : \mathbf{VA}^{\perp}) = \mathscr{C}(\mathbf{A} : \mathbf{V})$. □

We shall use the short notation

$$\mathbf{H} = \mathbf{P}_{\mathbf{X}}\,, \qquad \mathbf{M} = \mathbf{I}_n - \mathbf{H}\,,$$

and thereby the ordinary least squares estimator (OLSE) of $\mathbf{X}\boldsymbol{\beta}$ is $\mathbf{Hy}$; we will denote $\mathbf{Hy} = \mathbf{X}\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is any solution to $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$. If $\mathbf{X}$ has full column rank then $\boldsymbol{\beta}$ is estimable and its OLSE is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^{+}\mathbf{y}$.

Characterizing the equality of the OLSE and the BLUE of $\mathbf{X}\boldsymbol{\beta}$ has received a lot of attention in the statistical literature, the major breakthroughs being made by Rao [40], Zyskind [58], and Kruskal [27]; for a review, see Puntanen & Styan [37], and for some special remarks, Markiewicz, Puntanen & Styan [33], and O.M. Baksalary, Trenkler & Liski [9].

Theorem 6.3 gives immediately several equivalent characterizations for the OLSE and the BLUE to be equal, some of them are collected in Theorem 6.5. Notice that then the equality between OLSE and BLUE occurs with probability 1 but in what follows, we drop off the phrase "with probability 1".

Theorem 6.5. *Consider the general linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. *Then* $\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ *if and only if any one of the following five equivalent conditions holds:*
(a) $\mathbf{HV} = \mathbf{VH}$, (b) $\mathbf{HVM} = \mathbf{0}$, (c) $\mathscr{C}(\mathbf{VX}) \subset \mathscr{C}(\mathbf{X})$,
(d) $\mathscr{C}(\mathbf{X})$ *has a basis comprising a set of* $r = \mathrm{rank}(\mathbf{X})$ *orthonormal eigenvectors of* $\mathbf{V}$,
(e) $\mathbf{V} = a\mathbf{I}_n + \mathbf{HN}_1\mathbf{H} + \mathbf{MN}_2\mathbf{M}$ *for some* $a \in \mathbb{R}$, *and matrices* $\mathbf{N}_1$ *and* $\mathbf{N}_2$ *such that* $\mathbf{V}$ *is nonnegative definite.*

Using, for example, Rao & Mitra [46, p. 24] and Ben-Israel & Greville [10, p. 52], we obtain the following.

Theorem 6.6. *The general solution for* $\mathbf{G}$ *satisfying* $\mathbf{G}(\mathbf{X} : \mathbf{VX}^{\perp}) = (\mathbf{X} : \mathbf{0})$ *can be expressed, for example, in the following ways:*
(a) $\mathbf{G}_1 = (\mathbf{X} : \mathbf{0})(\mathbf{X} : \mathbf{VX}^{\perp})^{-} + \mathbf{F}_1\mathbf{Q}_{\mathbf{W}}$,
(b) $\mathbf{G}_2 = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-} + \mathbf{F}_2\mathbf{Q}_{\mathbf{W}}$,
(c) $\mathbf{G}_3 = \mathbf{I}_n - \mathbf{VX}^{\perp}[(\mathbf{X}^{\perp})'\mathbf{VX}^{\perp}]^{-}(\mathbf{X}^{\perp})' + \mathbf{F}_3\mathbf{Q}_{\mathbf{W}}$,
(d) $\mathbf{G}_4 = \mathbf{H} - \mathbf{HVX}^{\perp}[(\mathbf{X}^{\perp})'\mathbf{VX}^{\perp}]^{-}(\mathbf{X}^{\perp})' + \mathbf{F}_4\mathbf{Q}_{\mathbf{W}}$,
where $\mathbf{F}_1 \dots, \mathbf{F}_4$ *are arbitrary matrices,* $\mathbf{Q}_{\mathbf{W}} = \mathbf{I}_n - \mathbf{P}_{\mathbf{W}}$, *and* $\mathbf{W} \in \mathcal{W}$, *where* $\mathcal{W}$ *is defined as in* (17).

In view of the consistency condition (14), we have $\mathbf{y} \in \mathscr{C}(\mathbf{W})$ and hence the terms $\mathbf{F}_i\mathbf{Q}_{\mathbf{W}}\mathbf{y}$ disappear with probability 1. We observe, for example, that

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{Hy} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{My} = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{My},$$

or, denoting shortly $\mathbf{X}\tilde{\boldsymbol{\beta}} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ and $\mathbf{X}\hat{\boldsymbol{\beta}} = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta})$,

$$\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{My}.$$

It is easy to confirm that

$$\mathrm{cov}(\mathbf{X}\tilde{\boldsymbol{\beta}}) = \mathbf{HVH} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH} = \mathrm{cov}(\mathbf{X}\hat{\boldsymbol{\beta}}) - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH}. \tag{23}$$

When $\mathbf{X}$ has full column rank and $\mathbf{V}$ is positive definite, then $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ while the corresponding covariance matrices are

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{VX}(\mathbf{X}'\mathbf{X})^{-1}, \quad \mathrm{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{24}$$

On the other hand, in light of (23) we have

$$\mathrm{cov}(\tilde{\boldsymbol{\beta}}) = \mathrm{cov}(\hat{\boldsymbol{\beta}}) - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MVX}(\mathbf{X}'\mathbf{X})^{-1}. \tag{25}$$

It is interesting to note that in (25) the covariance matrix $\mathbf{V}$ need not be positive definite. If $\mathbf{V}$ is positive definite, then combining (24) and (25) yields the following:

Theorem 6.7. *Consider the linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, *where* $\mathbf{X}$ *has full column rank and* $\mathbf{V}$ *is positive definite. Then*

$$\begin{aligned}\mathrm{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} &= (\mathbf{X}'\mathbf{X})^{-1}[\mathbf{X}'\mathbf{VX} - \mathbf{X}'\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MVX}](\mathbf{X}'\mathbf{X})^{-1} \\ &= \mathrm{cov}(\hat{\boldsymbol{\beta}}) - \mathbf{X}^{+}\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV}(\mathbf{X}^{+})',\end{aligned} \tag{26}$$

and

$$\mathrm{cov}(\mathbf{X}\tilde{\boldsymbol{\beta}}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}' = \mathbf{HVH} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH}.$$

Among the first places where (26) occurs are the papers by Khatri [26, Lemma 1], Rao [40, Lemmas 2a, 2b, 2c], and Rao [43, p. 77]. Theorem 6.7 offers a convenient way to express the so-called Watson efficiency, see Watson [57, p. 330], as

$$\begin{aligned}\phi = \frac{|\mathrm{cov}(\tilde{\boldsymbol\beta})|}{|\mathrm{cov}(\hat{\boldsymbol\beta})|} &= \frac{|\mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{X}| \cdot |\mathbf{X}'\mathbf{X}|^{-2}}{|\mathbf{X}'\mathbf{V}\mathbf{X}| \cdot |\mathbf{X}'\mathbf{X}|^{-2}} = \frac{|\mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{X}|}{|\mathbf{X}'\mathbf{V}\mathbf{X}|}\\ &= |\mathbf{I}_p - (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{X}|\,.\end{aligned}$$

Above $|\cdot|$ refers to the determinant. For related considerations, see Puntanen, Styan & Isotalo [39, §10.7–10.8] and the references therein.

In this context we may briefly say a couple of words about the matrix product

$$\dot{\mathbf{M}} := \mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\,,$$

which appears in several formulas above. If $\mathbf{V}$ is positive definite and $\mathbf{V}^{1/2}$ is its positive definite symmetric square root, and $\mathbf{Z}$ is a matrix having full column rank with the property $\mathscr{C}(\mathbf{Z}) = \mathscr{C}(\mathbf{M})$, then we obviously have

$$\dot{\mathbf{M}} = \mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M} = \mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{Z}}\mathbf{V}^{-1/2} = \mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}',$$

which is clearly unique. In general, the matrix $\dot{\mathbf{M}}$ is not necessarily unique with respect to the choice of $(\mathbf{M}\mathbf{V}\mathbf{M})^{-}$. Moreover, for positive definite $\mathbf{V}$ we have

$$\dot{\mathbf{M}} = \mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M} = (\mathbf{M}\mathbf{V}\mathbf{M})^{+} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1},$$

and if $\mathbf{H}\mathbf{P}_{\mathbf{V}}\mathbf{M} = \mathbf{0}$ then, see Isotalo, Puntanen & Styan [24, Th. 2.1],

$$\mathbf{P}_{\mathbf{V}}\dot{\mathbf{M}}\mathbf{P}_{\mathbf{V}} = \mathbf{P}_{\mathbf{V}}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{P}_{\mathbf{V}} = \mathbf{V}^{+} - \mathbf{V}^{+}\mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+}.$$

The matrix $\dot{\mathbf{M}}$ is very handy in many connections related to linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol\beta, \mathbf{V}\}$. For example, the ordinary, unweighted sum of squares of errors SSE is defined as

$$\mathrm{SSE} = \mathrm{SSE}(\mathbf{I}) = \min_{\boldsymbol\beta}\|\mathbf{y} - \mathbf{X}\boldsymbol\beta\|^2 = \mathbf{y}'\mathbf{M}\mathbf{y}\,,$$

while the weighted SSE is (when $\mathbf{V}$ is positive definite)

$$\begin{aligned}\mathrm{SSE}(\mathbf{V}) &= \min_{\boldsymbol\beta}\|\mathbf{y} - \mathbf{X}\boldsymbol\beta\|^2_{\mathbf{V}^{-1}} = \|\mathbf{y} - \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y}\|^2_{\mathbf{V}^{-1}} = \mathbf{y}'[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{y}\\ &= \mathbf{y}'\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} = \mathbf{y}'\dot{\mathbf{M}}\mathbf{y}\,.\end{aligned}$$

In the general case, the weighted SSE can be defined as

$$\mathrm{SSE}(\mathbf{V}) = (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol\beta})'\mathbf{W}^{-}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol\beta})\,,$$

where $\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{X}'$, with $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$. Then, again,

$$\mathrm{SSE}(\mathbf{V}) = (\mathbf{y} - \mathbf{X}\tilde{\boldsymbol\beta})'\mathbf{W}^{-}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol\beta}) = \mathbf{y}'\dot{\mathbf{M}}\mathbf{y}\,.$$

For further properties of $\dot{\mathbf{M}}$, see Puntanen, Styan & Isotalo [39, Ch. 15] and Isotalo, Puntanen & Styan [24]. Some related considerations appear also in Markiewicz [32, pp. 415–416], LaMotte [28, pp. 323–324], and Searle, Casella & McCulloch [48, pp. 451–452].

What about if we require that *every* representation of the BLUE under $\mathscr{M}_1 = \{\mathbf{y}, \mathbf{X}\boldsymbol\beta, \mathbf{V}_1\}$ continues to be BLUE under $\mathscr{M}_2 = \{\mathbf{y}, \mathbf{X}\boldsymbol\beta, \mathbf{V}_2\}$? The answer is given in Theorem 6.8. For the proof and related discussion, see, e.g., J.K. Baksalary & Mathew [2, Th. 3], Mitra & Moore [35, Th. 4.1–4.2], Rao [41, Lemma 5], Rao [42, Th. 5.2, Th. 5.5], Rao [44, p. 289], Tian [52], Tian & Takane [54, 55], and Hauke, Markiewicz & Puntanen [21].

Theorem 6.8. *Consider the linear models $\mathscr{M}_1 = \{\mathbf{y}, \mathbf{X}\boldsymbol\beta, \mathbf{V}_1\}$ and $\mathscr{M}_2 = \{\mathbf{y}, \mathbf{X}\boldsymbol\beta, \mathbf{V}_2\}$. Then every representation of the* BLUE *for $\mathbf{X}\boldsymbol\beta$ under $\mathscr{M}_1$ remains the* BLUE *for $\mathbf{X}\boldsymbol\beta$ under $\mathscr{M}_2$ if and only if any of the following equivalent conditions hold:*

(a) $\mathscr{C}(\mathbf{V}_2\mathbf{X}^{\perp}) \subset \mathscr{C}(\mathbf{V}_1\mathbf{X}^{\perp})$,

(b) $\mathbf{V}_2 = a\mathbf{V}_1 + \mathbf{X}\mathbf{N}_1\mathbf{X}' + \mathbf{V}_1\mathbf{X}^{\perp}\mathbf{N}_2(\mathbf{X}^{\perp})'\mathbf{V}_1$, *for some $a \in \mathbb{R}$, and matrices $\mathbf{N}_1$ and $\mathbf{N}_2$ such that $\mathbf{V}_2$ is nonnegative definite.*

6.4 The reduced model

Let us consider the partitioned linear model $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{I}_n\}$, where $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ has full column rank, $\mathbf{X}_1 \in \mathbb{R}^{n\times p_1}$, $\mathbf{X}_2 \in \mathbb{R}^{n\times p_2}$, $p = p_1 + p_2$. In light of the projector decomposition (15), we have $\mathbf{H} = \mathbf{P}_{(\mathbf{X}_1 : \mathbf{X}_2)} = \mathbf{P}_{\mathbf{X}_1} + \mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}$, where $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_1}$ and thereby

$$\mathbf{H}\mathbf{y} = \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{P}_{\mathbf{X}_1}\mathbf{y} + \mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}\mathbf{y}\,. \tag{27}$$

Premultiplying (27) by $\mathbf{M}_1$ gives

$$\mathbf{M}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 = \mathbf{P}_{\mathbf{M}_1\mathbf{X}_2}\mathbf{y} = \mathbf{M}_1\mathbf{X}_2(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y}\,. \tag{28}$$

In view of (11), $\operatorname{rank}(\mathbf{M}_1\mathbf{X}_2) = \operatorname{rank}(\mathbf{X}_2) = p_2$, and hence the left-most $\mathbf{M}_1\mathbf{X}_2$ can be cancelled from (28) and thus we obtain

$$\hat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y} := \hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12}). \tag{29}$$

Premultiplying the model $\mathscr{M}_{12}$ by the orthogonal projector $\mathbf{M}_1$ yields the reduced model

$$\mathscr{M}_{12\cdot 1} = \{\mathbf{M}_1\mathbf{y},\ \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2,\ \mathbf{M}_1\}\,.$$

Taking a look at the models, we can immediately make an important conclusion: the OLS estimators of $\boldsymbol{\beta}_2$ under the models $\mathscr{M}_{12}$ and $\mathscr{M}_{12\cdot 1}$ coincide:

$$\hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12\cdot 1}) = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y}. \tag{30}$$

The equality (30) is the result that Davidson & MacKinnon [14, §2.4] call the Frisch–Waugh–Lovell theorem; see Frisch & Waugh [15], Lovell [29, 30].

Let us take a quick look at the more general case when the partitioned linear model is $\mathscr{M}_{12} = \{\mathbf{y},\ \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,\ \mathbf{V}\}$. Premultiplying $\mathscr{M}_{12}$ by the orthogonal projector $\mathbf{M}_1$ yields the reduced model

$$\mathscr{M}_{12\cdot 1} = \{\mathbf{M}_1\mathbf{y},\ \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2,\ \mathbf{M}_1\mathbf{V}\mathbf{M}_1\}\,.$$

What about the BLUE of $\mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2$ in the reduced model $\mathscr{M}_{12\cdot 1}$? Let us denote

$$\{\,\mathrm{BLUE}(\mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2 \mid \mathscr{M}_{12})\,\} = \{\,\mathbf{A}\mathbf{y} : \mathbf{A}\mathbf{y} \text{ is BLUE for } \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2\,\}\,.$$

Before proceeding we notice that $\mathbf{K}_2'\boldsymbol{\beta}_2$ is estimable under $\mathscr{M}_{12}$ if and only if there exists a matrix $\mathbf{L}$ such that $\mathbf{L}(\mathbf{X}_1 : \mathbf{X}_2) = (\mathbf{0} : \mathbf{K}_2)$, i.e., see Groß & Puntanen [17, Lemma 1],

$$\mathscr{C}(\mathbf{K}_2) \subset \mathscr{C}(\mathbf{X}_2'\mathbf{X}_1^{\perp}) = \mathscr{C}(\mathbf{X}_2'\mathbf{M}_1).$$

Moreover, it is easy to confirm that $\mathbf{K}_2'\boldsymbol{\beta}_2$ is estimable under $\mathscr{M}_{12}$ if and only if $\mathbf{K}_2'\boldsymbol{\beta}_2$ is estimable under $\mathscr{M}_{12\cdot 1}$. Then we can formulate the generalized Frisch–Waugh–Lovell theorem as follows; see, e.g., Groß & Puntanen [17, Th. 4].

Theorem 6.9. *Every representation of the* BLUE *of* $\mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2$ *under* $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{V}\}$ *remains the* BLUE *under* $\mathscr{M}_{12\cdot 1} = \{\mathbf{M}_1\mathbf{y},\ \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2,\ \mathbf{M}_1\mathbf{V}\mathbf{M}_1\}$ *and vice versa, i.e., the sets of the* BLUE*s coincide:*

$$\{\,\mathrm{BLUE}(\mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2 \mid \mathscr{M}_{12})\,\} = \{\,\mathrm{BLUE}(\mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2 \mid \mathscr{M}_{12\cdot 1})\,\}\,.$$

In other words: Let $\mathbf{K}_2'\boldsymbol{\beta}_2$ *be an arbitrary estimable parametric function under* $\mathscr{M}_{12}$. *Then every representation of the* BLUE *of* $\mathbf{K}_2'\boldsymbol{\beta}_2$ *under* $\mathscr{M}_{12}$ *remains the* BLUE *under* $\mathscr{M}_{12\cdot 1}$ *and vice versa.*

Let $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ have full column rank, and $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$, but $\mathbf{V}$ is possibly singular. Then it appears that corresponding to (29) we have

$$\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) = (\mathbf{X}_2'\dot{\mathbf{M}}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\dot{\mathbf{M}}_1\mathbf{y}\,,$$

where $\dot{\mathbf{M}}_1 = \mathbf{M}_1(\mathbf{M}_1\mathbf{V}\mathbf{M}_1)^{-}\mathbf{M}_1 = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-}\mathbf{X}_1'\mathbf{V}^{-1}$.

For further references related to the Frisch–Waugh–Lovell theorem, see for example, Bhimasankaram & Sengupta [12, Th. 6.1], Sengupta & Jammalamadaka [51, §7.10], and Groß & Puntanen [18].

6.5 Best linear unbiased predictor, BLUP

Let $\mathbf{y}_f$ denote a $q \times 1$ unobservable random vector containing new future observations. The new observations are assumed to follow the linear model $\mathbf{y}_f = \mathbf{X}_f\boldsymbol{\beta} + \boldsymbol{\varepsilon}_f$, where $\mathbf{X}_f$ is a known $q \times p$ matrix, $\boldsymbol{\beta}$ is the same vector of unknown parameters as in $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, and $\boldsymbol{\varepsilon}_f$ is a q-dimensional random error vector associated with new observations. Then

$$\mathrm{E}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_f\end{pmatrix} = \begin{pmatrix}\mathbf{X}\boldsymbol{\beta}\\ \mathbf{X}_f\boldsymbol{\beta}\end{pmatrix} = \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_f\end{pmatrix}\boldsymbol{\beta}, \quad \mathrm{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_f\end{pmatrix} = \begin{pmatrix}\mathbf{V} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}.$$

For brevity, we denote

$$\mathscr{M}_f = \left\{\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_f\end{pmatrix}, \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_f\end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix}\mathbf{V} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}\right\}. \tag{31}$$

The linear predictor $\mathbf{By}$ is said to be unbiased for $\mathbf{y}_f$ if $\mathrm{E}(\mathbf{y}_f - \mathbf{By}) = \mathbf{0}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$. This is equivalent to $\mathbf{BX} = \mathbf{X}_f$. Now a linear unbiased predictor $\mathbf{By}$ is the best linear unbiased predictor, BLUP, for $\mathbf{y}_f$, if the Löwner ordering $\mathrm{cov}(\mathbf{y}_f - \mathbf{By}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{y}_f - \mathbf{Fy})$ holds for all $\mathbf{F}$ such that $\mathbf{Fy}$ is an unbiased linear predictor for $\mathbf{y}_f$.

The following theorem characterizes the BLUP; see, e.g., Christensen [13, p. 294], and Isotalo & Puntanen [23, p. 1015].

Theorem 6.10. *Consider the linear model $\mathscr{M}_f$ (with new unobserved future observations), defined as in* (31)*, where $\mathscr{C}(\mathbf{X}_f') \subset \mathscr{C}(\mathbf{X}')$. The linear predictor* $\mathbf{By}$ *is the best linear unbiased predictor* (BLUP) *for* $\mathbf{y}_f$ *if and only if* $\mathbf{B}$ *satisfies the equation*

$$\mathbf{B}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X}_f : \mathbf{V}_{21}\mathbf{X}^{\perp}).$$

The linear mixed model $\mathscr{L}$, say, can be specified as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \text{i.e.,} \quad \mathscr{L} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}, \mathbf{D}, \mathbf{R}\}, \tag{32}$$

where $\boldsymbol{\beta}$ is a vector of fixed parameters and $\boldsymbol{\gamma}$ a vector of random ones, with the known covariance matrices $\mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{R}$ and $\mathrm{cov}(\boldsymbol{\gamma}) = \mathbf{D}$, and expectations $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{E}(\boldsymbol{\gamma}) = \mathbf{0}$. We assume that the random effect $\boldsymbol{\gamma}$ and error term $\boldsymbol{\varepsilon}$ are uncorrelated and thereby $\mathrm{cov}(\mathbf{y}) = \mathbf{ZDZ}' + \mathbf{R} = \boldsymbol{\Sigma}$, say. Taking $\boldsymbol{\gamma}$ as the "new observation" it is easy to conclude, in view of Theorem 6.10, that the following holds.

Theorem 6.11. *Consider the linear mixed model $\mathscr{L}$, defined as in* (32)*. The the linear predictor* $\mathbf{Ay}$ *is the* BLUP *of $\boldsymbol{\gamma}$ under the mixed model $\mathscr{L}$ if and only if*

$$\mathbf{A}(\mathbf{X} : \boldsymbol{\Sigma}\mathbf{X}^{\perp}) = (\mathbf{0} : \mathbf{DZ}'\mathbf{X}^{\perp}).$$

6.6 Stochastic restrictions

Let us consider the fixed effects partitioned model

$$\mathscr{F}: \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \mathrm{cov}(\mathbf{y}) = \mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{R},$$

where both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are fixed (but unknown) coefficients, and supplement $\mathscr{F}$ with the stochastic restrictions $\mathbf{y}_0 = \boldsymbol{\gamma} + \boldsymbol{\varepsilon}_0$, where $\mathrm{cov}(\boldsymbol{\varepsilon}_0) = \mathbf{D}$. This supplement can be expressed as the partitioned model:

$$\mathscr{F}_* = \{\mathbf{y}_*, \mathbf{X}_*\boldsymbol{\pi}, \mathbf{V}_*\} = \left\{\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_0\end{pmatrix}, \begin{pmatrix}\mathbf{X} & \mathbf{Z}\\ \mathbf{0} & \mathbf{I}_q\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\gamma}\end{pmatrix}, \begin{pmatrix}\mathbf{R} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{pmatrix}\right\}.$$

We will need the matrix $\mathbf{X}_*^{\perp}$ for which, according to part (b) of Theorem 3.2, one choice is $\left(\begin{smallmatrix}\mathbf{I}_n\\ -\mathbf{Z}'\end{smallmatrix}\right)\mathbf{M}$, where $\mathbf{M} = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}}$, and so we have

$$\mathbf{V}_*\mathbf{X}_*^{\perp} = \begin{pmatrix}\mathbf{R} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{pmatrix}\begin{pmatrix}\mathbf{I}_n\\ -\mathbf{Z}'\end{pmatrix}\mathbf{M} = \begin{pmatrix}\mathbf{RM}\\ -\mathbf{DZ}'\mathbf{M}\end{pmatrix}. \tag{33}$$

Now the estimator $\mathbf{B}\mathbf{y}_*$ is the BLUE for $\mathbf{X}_*\boldsymbol{\pi}$ under the model $\mathscr{F}_*$ if and only if $\mathbf{B}$ satisfies the equation

$$\mathbf{B}(\mathbf{X}_* : \mathbf{V}_*\mathbf{X}_*^{\perp}) = (\mathbf{X}_* : \mathbf{0})\,. \tag{34}$$

Substituting (33) into (34) yields

$$\begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{X} & \mathbf{Z} & \mathbf{R}\mathbf{M} \\ \mathbf{0} & \mathbf{I}_q & -\mathbf{D}\mathbf{Z}'\mathbf{M} \end{pmatrix} = \begin{pmatrix} \mathbf{X} & \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q & \mathbf{0} \end{pmatrix}. \tag{35}$$

Using (35) Haslett & Puntanen [20, Th. 1] show that all properties of BLUEs and BLUPs in mixed model $\mathscr{L}$ can be considered using the augmented model $\mathscr{F}_*$, where both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are fixed parameters. Using the connection between the mixed model $\mathscr{L}$ the augmented model $\mathscr{F}_*$, the following result follows from Theorem 6.8 immediately.

Theorem 6.12. *Consider two mixed models:* $\mathscr{L}_i = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}, \mathbf{D}_i, \mathbf{R}_i\}$, *and denote* $\boldsymbol{\Sigma}_i = \mathbf{Z}\mathbf{D}_i\mathbf{Z}' + \mathbf{R}_i$ *and* $\mathbf{V}_{*i} = \left(\begin{smallmatrix} \mathbf{R}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_i \end{smallmatrix}\right)$, $i = 1, 2$. *Then every representation of the* BLUE *for* $\mathbf{X}\boldsymbol{\beta}$ *under* $\mathscr{L}_1$ *remains the* BLUE *for* $\mathbf{X}\boldsymbol{\beta}$ *under* $\mathscr{L}_2$ *and every representation of the* BLUP *for* $\boldsymbol{\gamma}$ *under* $\mathscr{L}_1$ *remains the* BLUP *for* $\boldsymbol{\gamma}$ *under* $\mathscr{L}_2$ *if and only if any of the following equivalent conditions holds:*

(a) *Every representation of the* BLUE *for* $\mathbf{X}_*\boldsymbol{\pi}$ *under* $\mathscr{F}_{*1}$ *remains the* BLUE *for* $\mathbf{X}_*\boldsymbol{\pi}$ *under* $\mathscr{F}_{*2}$.

(b) $\mathscr{C}(\mathbf{V}_{*2}\mathbf{X}_*^{\perp}) \subset \mathscr{C}(\mathbf{V}_{*1}\mathbf{X}_*^{\perp})$.

(c) $\mathscr{C}\begin{pmatrix} \mathbf{R}_2\mathbf{M} \\ \mathbf{D}_2\mathbf{Z}'\mathbf{M} \end{pmatrix} \subset \mathscr{C}\begin{pmatrix} \mathbf{R}_1\mathbf{M} \\ \mathbf{D}_1\mathbf{Z}'\mathbf{M} \end{pmatrix}$.

(d) $\mathscr{C}\begin{pmatrix} \boldsymbol{\Sigma}_2\mathbf{M} \\ \mathbf{D}_2\mathbf{Z}'\mathbf{M} \end{pmatrix} \subset \mathscr{C}\begin{pmatrix} \boldsymbol{\Sigma}_1\mathbf{M} \\ \mathbf{D}_1\mathbf{Z}'\mathbf{M} \end{pmatrix}$.

(e) *The matrix* $\mathbf{V}_{*2}$ *can be expressed as*

$$\mathbf{V}_{*2} = a\mathbf{V}_{*1} + \mathbf{X}_*\mathbf{N}_1\mathbf{X}_*' + \mathbf{V}_{*1}\mathbf{X}_*^{\perp}\mathbf{N}_2(\mathbf{X}_*^{\perp})'\mathbf{V}_{*1}$$

for some $a \in \mathbb{R}$ *and matrices* $\mathbf{N}_1$ *and* $\mathbf{N}_2$ *such that* $\mathbf{V}_{*2}$ *is nonnegative definite.*

Acknowledgement: Thanks go to the referees for helpful remarks. Part of this research was done during the meeting of a Research Group on Mixed and Multivariate Models in the Mathematical Research and Conference Center, Będlewo, Poland, October 2013, supported by the Stefan Banach International Mathematical Center.

References

[1] Baksalary J.K., An elementary development of the equation characterizing best linear unbiased estimators, Linear Algebra Appl., 2004, 388, 3–6

[2] Baksalary J.K., Mathew T., Linear sufficiency and completeness in an incorrectly specified general Gauss–Markov model, Sankhyā A, 1986, 48, 169–180

[3] Baksalary J.K., Mathew T., Rank invariance criterion and its application to the unified theory of least squares, Linear Algebra Appl., 1990, 127, 393–401

[4] Baksalary J.K., Puntanen S., Styan G.P.H., A property of the dispersion matrix of the best linear unbiased estimator in the general Gauss–Markov model, Sankhyā A, 1990, 52, 279–296

[5] Baksalary J.K., Rao C.R., Markiewicz A., A study of the influence of the "natural restrictions" on estimation problems in the singular Gauss–Markov model, J. Statist. Plann. Inference, 1992, 31, 335–351

[6] Baksalary O.M., Trenkler G., A projector oriented approach to the best linear unbiased estimator, Statist. Papers, 2009, 50, 721–733

[7] Baksalary O.M., Trenkler G., Between OLSE and BLUE, Aust. N. Z. J. Stat., 2011, 53, 289–303

[8] Baksalary O.M., Trenkler G., Rank formulae from the perspective of orthogonal projectors, Linear Multilinear Algebra, 2011, 59, 607–625

[9] Baksalary O.M., Trenkler G., Liski E.P., Let us do the twist again. Statist. Papers, 2013, 54, 1109–1119

[10] Ben-Israel A., Greville T.N.E., Generalized inverses: theory and applications, 2nd Ed., Springer, New York, 2003

[11] Ben-Israel A., The Moore of the Moore–Penrose inverse, Electron. J. Linear Algebra, 9, 150–157, 2002

[12] Bhimasankaram P., Sengupta D., The linear zero functions approach to linear models, Sankhyā B, 1996, 58, 338–351

[13] Christensen R., Plane answers to complex questions: the theory of linear models, 4th Ed. Springer, New York, 2011
[14] Davidson R., MacKinnon J.G., Econometric theory and methods, Oxford University Press, New York, 2004
[15] Frisch R., Waugh F.V., Partial time regressions as compared with individual trends, Econometrica, 1933, 1, 387–401
[16] Groß J., The general Gauss–Markov model with possibly singular dispersion matrix, Statist. Papers, 2004, 45, 311–336
[17] Groß J., Puntanen S., Estimation under a general partitioned linear model, Linear Algebra Appl., 2000, 321, 131–144
[18] Groß J., Puntanen S., Extensions of the Frisch–Waugh–Lovell Theorem, *Discuss. Math. Probab. Stat.*, 2005, 25, 39–49
[19] Harville D.A., Matrix algebra from a statistician's perspective, Springer, New York, 1997
[20] Haslett S.J., Puntanen S., Equality of BLUEs or BLUPs under two linear models using stochastic restrictions, Statist. Papers, 2010, 51, 465–475
[21] Hauke J., Markiewicz A., Puntanen S., Comparing the BLUEs under two linear models, Comm. Statist. Theory Methods, 2012, 41, 2405–2418
[22] Herr D.G., On the history of the use of geometry in the general linear model, Amer. Statist., 1980, 34, 43–47
[23] Isotalo J., Puntanen S., Linear prediction sufficiency for new observations in the general Gauss–Markov model, Comm. Statist. Theory Methods, 2006, 35, 1011–1023
[24] Isotalo J., Puntanen S., Styan G.P.H., A useful matrix decomposition and its statistical applications in linear regression, Comm. Statist. Theory Methods, 2008, 37, 1436–1457
[25] Kala R., Projectors and linear estimation in general linear models, Comm. Statist. Theory Methods, 1981, 10, 849–873
[26] Khatri C.G., A note on a MANOVA model applied to problems in growth curves, Ann. Inst. Statist. Math., 1966, 18, 75–86
[27] Kruskal W., When are Gauss–Markov and least squares estimators identical? A coordinate-free approach, Ann. Math. Statist., 1968, 39, 70–75
[28] LaMotte L.R., A direct derivation of the REML likelihood function, Statist. Papers, 2007, 48, 321–327
[29] Lovell M.C., Seasonal adjustment of economic time series and multiple regression analysis, J. Amer. Statist. Assoc., 1963, 58, 993–1010
[30] Lovell M.C., A simple proof of the FWL Theorem, J. Econ. Educ., 2008, 39, 88–91
[31] Margolis M.S., Perpendicular projections and elementary statistics, Amer. Statist., 1979, 33, 131–135
[32] Markiewicz A., On dependence structures preserving optimality, Statist. Probab. Lett., 2001, 53, 415–419
[33] Markiewicz A., Puntanen S., Styan G.P.H., A note on the interpretation of the equality of OLSE and BLUE, Pakistan J. Statist., 2010, 26, 127–134
[34] Marsaglia G., Styan G.P.H., Equalities and inequalities for ranks of matrices, Linear Multilinear Algebra, 1974, 2, 269–292
[35] Mitra S.K., Moore B.J., Gauss–Markov estimation with an incorrect dispersion matrix, Sankhyā A, 1973, 35, 139–152
[36] Mitra S.K., Rao C.R., Projections under seminorms and generalized Moore–Penrose inverses, Linear Algebra Appl., 1974, 9, 155–167
[37] Puntanen S., Styan G.P.H., The equality of the ordinary least squares estimator and the best linear unbiased estimator (with discussion), Amer. Statist., 1989, 43, 151–161 [Commented by O. Kempthorne on pp. 161–162 and by S.R. Searle on pp. 162–163, Reply by the authors on p. 164]
[38] Puntanen S., Styan G.P.H., Reply [to R. Christensen (1990), R.W. Farebrother (1990), and D.A. Harville (1990)] (Letter to the Editor), Amer. Statist., 1990, 44, 192–193
[39] Puntanen S., Styan G.P.H., Isotalo J., Matrix tricks for linear statistical models: our personal top twenty, Springer, Heidelberg, 2011
[40] Rao C.R., Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Berkeley, California, 1965/1966, vol. 1, L.M. Le Cam and J. Neyman, eds., University of California Press, Berkeley, 355–372, 1967
[41] Rao C.R., A note on a previous lemma in the theory of least squares and some further results, Sankhyā A, 1968, 30, 259–266
[42] Rao C.R., Unified theory of linear estimation, Sankhyā A 1971, 33, 371–394. [Corrigendum (1972), 34, p. 194 and p. 477]
[43] Rao C.R., Linear statistical inference and its applications, 2nd Ed., Wiley, New York, 1973
[44] Rao C.R., Representations of best linear unbiased estimators in the Gauss–Markoff model with a singular dispersion matrix, J. Multivariate Anal., 1973, 3, 276–292
[45] Rao C.R., Projectors, generalized inverses and the BLUE's, J. R. Stat. Soc. Ser. B Stat. Methodol., 1974, 336, 442–448
[46] Rao C.R., Mitra S.K., Generalized nverse of matrices and its applications, Wiley, New York, 1971
[47] Rao C.R., Rao M.B., Matrix algebra and its applications to statistics and econometrics, World Scientific, River Edge, NJ, 1998
[48] Searle S.R., Casella G., McCulloch C.E., Variance components, Wiley, New York, 1992
[49] Seber G.A.F., The linear hypothesis: a general theory, 2nd Ed., Griffin, London, 1980
[50] Seber G.A.F., Lee A.J., Linear regression analysis, 2nd Ed. Wiley, New York, 2003
[51] Sengupta D., Jammalamadaka S.R., Linear models: an integrated approach, World Scientific, River Edge, NJ., 2003
[52] Tian Y., On equalities for BLUEs under misspecified Gauss–Markov models, Acta Math. Sin. (Engl. Ser.), 2009, 25, 1907–1920
[53] Tian Y., Beisiegel, M., Dagenais E., Haines C., On the natural restrictions in the singular Gauss–Markov model, Statist. Papers, 2008, 49, 553–564
[54] Tian Y., Takane Y., Some properties of projectors associated with the WLSE under a general linear model. J. Multivariate Anal., 2008, 99, 1070–1082
[55] Tian Y., Takane Y., On V-orthogonal projectors associated with a semi-norm, Ann. Inst. Statist. Math., 2009, 61, 517–530

[56] Trenkler G., On the singularity of the sample covariance matrix, J. Stat. Comput. Simul., 1995, 52, 172–173
[57] Watson G.S., Serial correlation in regression analysis, I, Biometrika, 1955, 42, 327–341
[58] Zyskind G., On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models, Ann. Math. Statist., 1967, 38, 1092–1109
[59] Zyskind G., Martin F.B., On best linear estimation and general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure, SIAM J. Appl. Math., 1969, 17, 1190–1202